

Europäisches Patentamt

European Patent Office

Office européen des brevets

(1) Publication number:

0 079 739

A2

12

# EUROPEAN PATENT APPLICATION

(2) Application number: 82305926.6

(22) Date of filing: 08.11.82

(5) Int. Cl.<sup>3</sup>: C 12 N 15/00

C 12 N 1/00, C 12 P 21/02 C 07 H 21/04, C 07 C 103/52

//C12R1/19, C12R1/865

30 Priority: 12.11.81 US 320632

Date of publication of application: 25.05.83 Bulletin 83/21

(84) Designated Contracting States: BE CH DE FR GB IT LI NL SE (1) Applicant: THE UPJOHN COMPANY 301 Henrietto Street Kolamazoo, Michigan 49001(US)

(2) Inventor: Dugalczyk, Achilles e/o The Upjohn Company 301 Henrietta Street Kalamazoo Michigan 49001(US)

(4) Representative: Perry, Robert Edward ot al, GILL JENNINGS & EVERY 53-64 Chancery Lane London WC2A 1HN(GB)

(54) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

(5) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

Ø 688 ≥

000

<u>@</u>

UPJOHN 1 GJE 70/2056/02

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in

5 development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly a-fetoprotein, but the synthesis decreases drastically after birth. Recently,

10 Law et al determined the complete sequence of mouse

α-fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been 15 reached from studies on the α-fetoprotein genes of the

rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum

20 mRNA has been determined from recombinant cDNA clones and
from a primer-extended cDNA synthesis on the mRNA
template. The sequence comprises 2,078 nucleotides,
starting upstream of a potential ribosome binding site
in the 5'-untranslated region. It contains all the

- 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal
- 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino
- 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

### DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleo-25 tides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Mucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table  $^{30}$  1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T 
$$T^{C}$$
T C T T C T G T......albumin mRNA (3')...G A G G A A G G C G U C C  $m_{2}^{C}$ A  $m_{2}^{C}$ A......185 RNA

35

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since prepeptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-13621. This is the first report of the presence and the sequence of a prepeptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the propeptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the concensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

# TABLE 1

		(30)	(170)	(190)	(350)	(044)	(330)	(620)	(710)	(300)
	5	AGC	20 1ys AAA (	50 818 GCA (	80 leu CTT (	110 Pro CCA	140 try TAT	170 gln CAA	200 cys TCT	230 glu GAA (
	5	phe :	phe TTC	phe TTT	th. ACT	AAC	leu TTA	169 cys TG:		ala GCA
		leu i	AAT	alu CAA	e Le GCA	ase	tyr TAC	168 cys TGT	leu CTC	
		phe	GAA	thr	val	CAT	Tys AAA	5 A8	ser ala lys qin arq leu lys TCT GCC AAA CAG AGA CTC AAG	gin arg phe pro lys ala glu phe CAG AGA TTT CCC AAA GCT GAG TTT
		leu crr	g lu	val	thr	1 ya	1ys AAA	thr	cag Cag	ala GCT
	10	-10 ser leu leu TCC CTT CTT		glu val GAA GTA	27 203 100	leu qin his Jys TTG CAA CAC AAA	leu TTG	phe TTT	1ys AAA	1ys AAA
		ser TCC	leu aly TTG GGA		leu TTA	g In	phe TTT	ala GCT	818 600	200
			GAT	val GTG	1ys AAA	leu gin TTG CAA	thr	ala ala ccr ccr	ser TCT	phe TTT
		p r o trp val tlu phe lle TGC GTA ACC TTT ATT	1ya AAA	40 his val lys leu val asn CAT GTA AAA TTA GTG AAT	70 1eu phe gly asp lys leu cys CTT TTT CGA GAC AAA TTA TCC	phe	q 1 c	1ys AAA		arg AGA
		p r o trp val tlu phe TGC GTA ACC TTT	phe	lys AAA	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	101 9ys 1GC	alu GAA	eyf Tat	ala CCT	gla
,	15	p r val t	10 11 12 13	40 Val lys GTA AAA	70 phe TTT	충축용	130 asn AAT	160 279 AGG	190 1ys	220 leu ser CTG AGC
		5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	h i s CAT	nts CAT	leu CTT	asn AAT	asa GAC	A A	41 666	
			ala GCT	esp GAT	thr	<b>P S</b>	h13 CAT	ala CCT	eg g	ala arg GCT CGC
	•	-18 Met lys ATG AAG	val	eg 45	h13 CAT	91.9 CCC	phe TTT	phe TTT	asp GAT	val ala GTA CCT
	20	-18 Met 1ys GCTTTTCTCTTGTCACCCCACACCCCTTTGCCACA ATG AAG	g g	phe 777	62 cys asp lys ser leu his TGT GAC AAA TCA CTT CAT	glu pro GAA CCT	124 oys thr sls phe IGC ACT GCT TTT	Phe TTC	glu leu arg d GAA CTF CGG (	
	20	166	ger AGT	e S	ser ICA	916 GAA	thr	leu CTT	leu CT7	trp ala TCC CCA
		CCT1	lys AAG	% 8y3 TGT	1ys AAA	gla	124 9y8 1GC	leu CTC	<b>5</b> 89	5 5
		rage	his CAC	cAG	asp CAC	ela lys CCA AAA	met ATG	age Ge	asp GAT	lys sla AAA GCA
		/222	ala CCA	gln	62 cys TGT		val CTG	Pro CCG	leu CTC	1ya AA
	25	CAAC	-1 1 arg asp CGA GAT	30 tyr leu TAT CTT	60 glu asn GAA AAT	91 Cys	asp CAT	150 tyr ala pro TAT GCC CCG	180 pro lys CCA AAG	210 ala phe GCT TTC
		เราต		oc tyr TAT	8 5 §	8 8 100	120 val CTT			
		:TCT1	r o il phe arg ic TTT CGT	gln	ala GCT	asp CAC	glu GAG	phe TTT	leu TTG	glu arg GAA AGA
		TI.	e by the time that the time the time that the time the time the time the time the time the time that the time the time that the	ela CCT	ser TCA	ala CCT	pro CCA	tyr TAC	leu CTG	glu
		និ	7 val crc	phe TTT	glu	glu met GAA ATG	val arg	pro CCT	177 ala cys GCC TGC	gly ccA
	30		ely cct	စ္ခါ စီ ၁၁၁	asp GAT	glu me GAA AT	leu val TTG GTG	h1s CAT	၁၁၁	g TT
			arg AGC	11e Att	ala GCT	thr tyr gly ACC TAT GGT	leu 11G	arg AGA	ala CCT	lys AAA
			-1 ser 700	leu TTG	val CTT	tyr IAT	arg CGA	85 85	PAS AS	gln CAA
			-1 ala tyr ser GCT TAT TCC	leu val leu tle TTG GTG TTG ATT	53 cys TGT	glu thr tyr gly GAA ACC TAT GGT	leu pro arg CTC CCC CGA	ile ala arg arg his pro ATT CCC AGA AGA CAT CCT	ala asp lys c	ser leu gin lys phe AGT CTC CAA AAA TTT
	35		-1 -6 p ser ala tyr ser arg gly va TCC CCT TAT TCC ACC GGT GT	leu TTG	ACA ACA			11e ATT	ala GCT	ser
			3e <b>r</b> 1CG	21 ala ccc	178 AAA	arg CCT	111 asn AAC	141 GAA	171 ala GCT	201 ala GCC

	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
_	260 leu CTT	290 11e ATT	320 ala Gri	350 ala GCC	380 Jeu CCT	410 ara CGT	&&O h1s CAT	&70 ser AGT	1 ys	530 val GTT
5	asp GAC	289 cys TGC	tyr		pro CCT	val	1ys AAA	val GTA	0 D	leu CTT
	ala CC	h is CAC	asn AAC	8 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	lys AAA	leu TTA	438 cys TGT	5 S	val GTT	ala GCA
	arg AGG	ser TCC	316 asp val cys lys asn CAT GTT TGC AAA AAC	leu CTG	phe TTT	ala leu leu GCG CTG TTA	437 cys TGT	thr	asp glu thr tyr val GAT GAA ACA TAC GTT	ain ile lys lys ain thr CAA ATC AAG AAA CAA ACT
	asp GAC		316 cys TGC	val leu leu GTG CTG CTG	GAA CAA		lys AAA	1ys AAA	thr	aln CAA
10	asp CAT	leu glu TTG CAA	asp val CAT GTT	leu CTG	phe asp TTC CAT	asn	ser AGC	a cac	olu CAA	1ys AAA
	ele Cr	leu TTG	asp CAT	val CTG		£ 95		his CAT	asp CAT	lys lys aln AAG AAA CAA
	253 cys TGT	lya pro leu leu glu lys AAA CCT CTG TTG GAA AAA	1ys AAG	vál GTC	ala iys val GCC AAA GTG	phe TTC	430 asn leu qly lys val AAC CTA GGA AAA GTG	leu TTG	490 ala leu glu val GCT CTG GAA GTC	11e ATC
	ale SA	pro CCT	ser AGT	ser TCT	lys AAA	400 glu tyr lys GAG TAC AAA	Jys AAA	val GTG	g)u GAA	arg aln AGA CAA
•		1ys AAA		tyr TAC	هاه در	tyr TAC	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5		leu CTG	514 cys thr leu ser qlu lys qlu arq TCC ACA CTT TCT GAG AAG GAG AGA
15	250 asp leu leu GAT CTG CTT	280 41u GAA	310 phe val TTT CTT	340 8sp GAT	370 tyr TAT	400 41 u	&30 asn leu AAC CTA	460 gin leu CAG TTA	490 ser ala TCA GCT	520 ser qlu lys qlu TCT GAG AAG GAG
		279 cys TGT	phe TTT	pro CCT	369 alu cys CAA TGC	<u> </u>	asa			1ys AAG
	nis gly cat gga	278 91u ays GAA TGC	asp CAT	his CAT	5 A	leu CTT	ser arg ICA AGA	leu asn CTG AAC	cys phe TGC TTT	a Ju CAG
	FIS	<b>a</b> 8	ala GCT	arg his p AGG CAT (	M Is	aln CAG		اور 150	cys 160	ser TCT
20	245 246 thr glu cys cys ACG GAA TGC TGC	ser lys leu lys AGT AAA CTG AAG	ser leu ala TCA TTA GCT	AG AG	pro CCT	cy of cy	glu val CAC GTC	val	pro CCA	leu CTT
20	285 Cys 100	leu CTG	leu TTA	ala GCA	asp GAT	phe TTT		va) CTG	5 43 5 43	thr ACA
	glu	ser lys AGT AAA	ser TCA	glu tyr GAA TAT	ala GCA	glu leu GAG CTT	val GTA	ser Ser	arg	514 11e cys ATA TGC
	thr ACG		pro CCT		ala GCT		pro thr leu val CCA ACT CTT GTA	tyr leu TAI CTA	480 ser leu val asn arg TCC TTG GTG AAC AGG	11e ATA
	n Is	ser TCC	leu TTG	tyr TAT	ala CCC	392 cys TGT	thr		val GTG	asp
25	240 1ys val AAA GTC	11e ATC	asp	leu TTG	361 cys TGT	asn AAT	ere CCA	450 glu asp GAA GAC	leu TTG	510 his ala asp cat GCA CAT
	240 1 ys AAA	270 ser TCG	300 818 GCT	330 phe TTT	360 cys TGC	390 gln CAA	420 thr		480 ser TCC	
	thr	asp	85 CCT	met ATG	glu lys	ile lys ATC AAA	val ser GTC TCA	448 oys ela TGT GCA	glu GAA	thr phe ACC TTC
	1 E	gla	met ATG	917	glu GAG	- 4		448 0ys 1GT	thr ACA	
	asp GAT	glu asn GAA AAT	a)u GAG	leu TTG	leu CTA	leu TTA	gIn CAA	နို့ ဗ	477 cys TGC	asn ala glu thr phe AAT GCT GAA ACA TTC
30	lys leu val thr AAG TTA GTG ACA	g &	asp CAT	val phe GTC TTC	thr	gin asn CAG AAT	e d	lys arg æet pro AAA AGA ATG CCC	476 cys TCC	thr ACA
	val GTG	265 11e cys ATC TGT	glu asn GAA AAT	val GTC	thr ACC	gla CAG	val GTA	arg AGA	1 ys AAA	g &
	leu TTA	11e ATC	glu GA	asp GAT	ag Se	glu pro GAG CCT	1ys	lys Aaa	thr ACC	ala CCT
	1ys AAG	lys tyr AAG TAT	val GTG	ala lys asp val phe GCA AAG GAT GTC TTC	tyr TAT	glu glu pro GAA GAG CCT	thr lys lys val pro gln ACC AAG AAA GTA CCC CAA	glu ala lys arg met pro GAA GCA AAA AGA ATG CCC	476 arg val thr lys cys AGA CTC ACC AAA TGC	asn AAT
	ser 700	1ys AAG		ala GCA	351 1ys thr tyr glu thr thr AAG ACA TAT GAA ACC ACT	91. 88	thr Acc	of S	arg ACA	501 glu phe CAG TTT
35	231 val GTT	261 ala GCC	291 ala GCC	321 g1u GAG	351 1ys AAG	381 val GTG	411 tyr TAC	4 5 TO	471 asp CAC	501 91u GAG

	(1790)	1883)	2002)
	560 1ys AAG	570 phe ala glu qlu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC CAG GAC GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAAG (1883)	ter ter ter ter ter sammen en sammen en sammen en sammen en sammen en sammen sammen sammen (2002)
5	559 cys 1Gr	ter ITAA	ICT
J	558 cys TGC	ICAT	E S
	lys AG	ATC/	ATA
	558 559 pro lys sla thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys cys ccc aag cca aca aca aaa cac caa ctc aaa cct at at cat c	rer IAA (	JAAAC
	val o	TA T	CTAA
10	eg E	7, 28	161
	ala nct	leu o	ACC
	SCT O	[ <sub>8</sub> ] 8	Š,
	oke 7	sla s	¥ ¥
	asp g	580 91n CAA	CTGT/
15	isp CAT	der o	STTG
	aet ATG	ela SCA	TTC
	val e	ela CCT	Terr
	ele CCT	CTT	TTT
20	1 ys	Jeu	rere
20	leu CTG	lys AA	TTCA
	91n CAA	1 ys	CTTA
	glu	el y	AAAG
	lys AAA	5 5	ATCA
25	Sto thr ACA	570 91c GAG	CAAC
	e e GCA	ele GC	AAAT
	1ys AAG	phe TTT	AACA
	<u>د</u> کی	567 9ys TCC	9282
	lys AAG	glu thr GAG ACC	ter
30	A13 CAC	g)u GAG	ter TCACA
	1 ys	1ys AAG	CCAT
	val lys	asp lys GAT AAG	41.J.
	leu CTC	asp GAC	ידראנ
35	531 91u GAG	561 ala GCT	247 2103

ICATTITICCCICITITICICTGIGCTICAATIAAIAAAAIGGAAAGAAICIAA..... 20 .....AA (2078)

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

#### 5 Example 1 Isolation of Messenger RNA

•

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 15 680-685.

#### Cloning Procedures Example 2

30

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, 25 S., et al., <a>Ibid.</a>]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [32p]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from <u>E. coli</u> RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this The E. coli RR1 hosts were lysed and then centrifuged to art. 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HR101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl<sub>2</sub>. Bacteria are then concentrated to one-tenth of this volume in CaCl<sub>2</sub> and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 μg/ml tetracycline) with 200 μl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

# 15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Rethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

## Example 4 DNA Sequencing

35

phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and Y[32p]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

#### Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable  $\underline{E}$ ,  $\underline{coli}$  host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

HB101(pHA206) - NRRL B-12550

5

 $\underline{\text{E. coli}}$  HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid. 20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

# Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straighforward task of restriction enzymology. There is only one MspI site in the overlapping
DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI
digestion of the two DNAs, followed by (ii) the use of ligase, an
enzyme that seals DNA fragments, will give the desired product.
Although two other undesired DNA species will also be obtained in the
course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA
species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

(a) Escherichia coli

- (b) Saccharomyces cerevisiae
- (a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the 5 existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoR1 DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one 10 of the yeast plasmid vectors, e.g., YEp6, at the Eco R1 cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

# 15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the begin-20 ning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been 25 documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466. Example 8 Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an <u>in situ</u> lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

## CLAIMS

- 1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.
- Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.
- 3. E. coli HB101 (pHA36) having the deposit accession number 10 NRRL B-12551.
  - 4. E. coli HB101 (pHA206) having the deposit accession number NRRL B-12550.
- 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

	<b>-</b>	=	=	=	<b>5</b>	6	5	<b>~</b>	· =
	(30)	(170)	(092)	(350)	(440)	(330)	(629)	(710)	230 qlu GAA (300)
	ACC	8 <del>1</del> 8	50 818 GCA	86 CT CT	110 pro CCA	140 try TAT	170 gla GAA	200 cys TGT	
5	ohe III	asn phe AAT TTC	phe TTT	thr	asa AAC	leu TTA	169 Cy3 TGC	lys AG	ه ا ه در ۸
J	leu ohe CTC TTT		glu CAA	8 50 8 4 20	aso GAC	tyr TAC	168 Cys TGT	leu CTC	phe TTT
	pha TTT	GAA	thr ACT	val GTT	asp GAT	1ys AAA	3 8	aln ara CAG AGA	ala qlu CCT CAG
	leu CTT	SA AS		thr ACA	1ys	1ys AAA	thr Aca		ala SCT
	0.	<del>2</del> 8	16. AS	75 Cys TCC	gin his CAA CAC	leu TTG	phe	ele lys CCC AAA	lys AAA
10		leu TTG	asn	leu TTA	ala CAA	phe TTT	ala CCT		500
		asp GAT	val GTG	asp lys CAC AAA	leu TTG	qlu thr GAG ACA	ala GCT	10°	phe TTT
		Bys AAA	leu TTA		phe TTC	58	1ys	16G	ACA
		phe 777	40 val lys gta aaa	45 56	100 101 qlu cys CAA TCC	£ 8	160 arg tyr AGC TAT	190 Iys ele AAG GCT	220 arg leu ser gln CGC CTG AGC CAG
		S F F F F F F F F F F F F F F F F F F F		70 phe TTT	9 5 AS	130 asn AAT		190 175 AAG	220 ser AGC
15		his CAT	asp his Cat Cat	thr leu ACC CTT	AAT	esp GAC	ala lys CCT AAA	alu aly GAA GGG	1ee CTG
		ele GCT	asp GAT	thr	ACA ACA	phe his TTT CAT		5 8 8	13 20 20
			34 gin ays pro phe glu CAG IGT CCA TTT GAA	asp lys ser leu his thr leu GAC AAA TCA CTT CAT ACC CTT	<del>7</del> 000	ala phe GCT TTT	phe ohe TTC TTT	esp GAT	ala GCT
		glu val GAG CTT	phe	leu CTT	gin glu pro CAA GAA CCT	ele CCT	phe TTC	glu leu org GAA CTT CGG	ela vel CCA GTA
20		ser	pro CCA	ser TCA	glu GAA	thr ACT	leu CTT	leu	ele CCA
		1ys AAG	3 69 E	lys ser AAA TCA	gla CAA	124 679 100	glu leu leu GAA CTC CTT	glu GAA	trp 766
		NI S CAC	ole CAG	asp GAC	1 ys	met ATG	ale GAA	esp CAT	<b>25</b>
		ala his lys ser glu val GCA CAC AAG AGT GAG GTT	9In CAG	62 cys TGT	= 8 8 8	val CTG	979 CCG	lys leu AAG CTC	phe lys ala trp TTC AAA GCA TGC
		asp CAT	leu	AAT	91 Cy3 TGT		ရက လ	lys AAG	phe TTC
25		- 55 55 55 55	30 tyr 1Af	66 91.8 683	8 8 9 3 3 3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	120 val GTT	150 tyr TAT	160 Pro CCA	210 ala GCT
		arg CGT	g] CAG	sla GCT	asp GAC	alu GAG	phe TTT	leu TTG	AS AS
		phe TTT	ala CCT	ser TCA	ala CCT	<b>2</b> 25	tyr TAC	leu CTG	3
		و روا درد	phe TTT	g p GAG	glu met GAA ATG	arg AGA	his pro CAT CCT	177 ala cys CCC TCC	91. GCA
20		g by GGT	ala phe CCC TTT	esp GAT	glu met GAA ATG	val CTG			phe TTT
30		arg ACC	11e ATT	ala CCT	aly cct	arg leu CGA TTG	arg AGA	lys ele AAA GCT	1ys Aaa
		el ser TCC		val CTT	tyr TAT		arg AGA		<b>5</b> 8
		tyr TAT	leu val TTG GTG	53 0ys TGT	glu thr GAA ACC	leu pro CTC CCC	- a 2	ga TAT	ser leu gln lys phe gly AGT CTC CAA AAA TTT GGA
		ala CCT	leu TTG	<b>ch</b>	<b>2</b> 8		11e ATT	ele CC	
35	<b>;</b>	ser 100	21° 81° CCC	173 AAA	arg CGT	111 asn AAC	age SA	171 ele CCT	201 818 CCC

	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
5	260 asp leu gac ett	290 11e ATT	320 ala GCT	350 ala GCC	380 leu CCT	&10 &rq CGT	440 his cat	470 ser AGT	5no 1ys AAA	530 val CTT (
	BSP	289 cys 1CC	tyr Tat	leu CTT	pro CCT	val	1ys	val	bro CCC	leu CTT
	818 800	ser his ICC CAC	asn	leu ara leu CTG AGA CTT	glu phe 1ys GAA TTT AAA	leu TTA	438 cys TGT	pro val CCA GTA	val   CTT (	ala leu GCA CTT
:	org Agg	ser Tor	316 val oya lya CTT TGC AAA	leu CTG	phe TTT	leu CTG	437 cys TGT		tyr TAC	
	asp asp GAT CAC	1ys AAA	316 9ys 1GC	leu CTG	g) u	al a	173 AAA	Tys AAA	thr ACA	oln thr CAA ACT
10		leu leu glu CTG TTG GAA	asp val CAT CTT	leu CTG	asb CAT		ACC	leu his qiu lys thr TTG CAT GAG AAA ACG	aju thr tyr GAA ACA TAC	<b>7</b> 8
	253 cys ols for cer	leu TTG	asp CAT	va) GTG	phe TTC	aln asn CAG AAT	414	ils c	asp cAT	ys B
	253 cys 161	leu CTG	1ys AAG			phe	val	leu his TTG CAT		le l
	og S	pro CCT	glu ser lys GAA AGT AAG	ser val TCT GTC	lys val AAA GTG	1ys AAA	lys AAA	vel	glu v	ain ile iys iys Caa atc aag aaa
15	250 leu leu glu CTG CTT GAA	T XS	age GAA	tyr	ala ငင	ty.	430 leu qly lys val qly ser CTA GGA AAA GTG GGC AGC	461 0ys TGT (	490 ala leu glu val GCT CTG GAA GTC	520 glu arq qin ile iys iya GAG AGA CAA ATC AAG AAA
15	250 1eu CTG	280 91u GAA	310 val CTT	340 88p GAT	370 tyr TAT	\$00 glu tyr GAG TAC	\$30 leu CTA	460 461 leu oys TTA TGT	490 ala ) GCT C	520 głu arq CAC ACA
	asp CAT	279 0ys TCT	phe TTT	Pro CCT	369 cys 7GC	56 S		gla J		Ys y
	914 664	278 cys TCC	GAT	his Cat	369 qlu cys CAA TGC	leu CTT	erg esn AGA AAC	asa g	phe s	glu lys CAG AAG
	nla CAT	# 8 8	ala GCT	Pre AGC	h s CAT	q Ja CAG	ser arg TCA AGA	leu asn CTG AAC	\$ 5 2 5	ser q
20	245 246 cys cys his TGC TGC CAT	ser lys leu lys glu AGT AAA CTG AAG GAA	ala GCT	arg AGA	pro CCT	g c		val GTC	pro cys phe ser CCA TGC TTT TCA	514 cys thr leu ser ICC ACA CTT ICT
		leu CTG	leu TTA	ala GCA	asp CAT	phe TTT	glu val GAG GTC	val GTG	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	14 S
	g Ju GAA	ser lys AGT AAA	ser TCA	tyr TAT	ala GCA	leu		ser 700	arg a	514 cys r 100 A
	vel his thr glu GTC CAC ACG GAA	ser AGT	pro CCT	g Pa	ala GCT	glu GAG	leu val CTT GTA	leu CTA	asn a	
	h 1s CAC	ser 700	1eu 77G	tyr	8 8 8 000	392 cys TGT	thr ACT	tyr TAT	val a	ele esp Ile GCA GAT ATA
25		11e ATC	asp CAC	leu TTG	361 0ys TGT	asn	pro CCA	asp GAC	480 ser leu val TCC TTG CTG	818 675
	240 178 AAA	270 ser TCG	300 a1a CCT	330 phe TTT	360 933 767	330 gln CAA	the thr ACT	450 91u GAA	480 ser 1	510 his (
	thr Acc	gin asp CAA CAT	pro cc1	met ATG	1ys AAG	1ys	ser TCA	919 CCA	91u 88	phe t
	leu CTT	ala CAA	met ATG	phe leu gly i	glu thr thr leu glu GAA ACC ACT CTA GAG	11e ATC	val	448 oys TCT		
30	leu val the asp le ITA GTG ACA GAT CT	glu asn GAA AAT	16 68	leu 11G	leu CTA	esn leu ile Aat :rta atc	gin val CAA GTG	6 TG 200	477 cys thr TGC ACA	ala glu thr phe thr GCT GAA ACA TTC ACC
	thr ACA	glu GAA	asp GAT	phe TTC	thr	asn	. ord	Met A	476 0ys TGC	4 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	val GTG	269 cys TGT	asn	val GTC	thr	pro gin		arg AcA	476 thr lys cys ACC AAA TGC	glu thr GAA ACA
	leu 11A	11e ATC	<b>₹</b>	asp GAT	914	pro	1ys AAA	1ys	thr 1	ola SCT O
	ser lys ICC AAG	Eyr	glu val GAA GTG	1ys AAG	tyr TAT	glu	1ys AAG	ele cca	val crc /	asn ala AAT GCT
35		Lys		al a GCA	th. ACA	glu GAA	thr lys lys val ACC AAG AAA GTA	glu c	ACA A	he a
	231 val GTT	261 ala ccc	291 818 GCC	321 91u CAG	351 Lys AAG	381 val GTG	tyr tyr TAC	441 Pro 9 CCT (	471 asp a	501 glu phe GAG TTT

	(1790)	leu ter tar (1883)	AGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTTCGTTGGTGTAAACCCAACACCCTGTCTAAAAACATAAATTTTTTAA (2002)
	560 173 AAG	ç X	<u> </u>
-	559 757	ter T&	E
5	558 cys TGC	ACAT	MIT
-	1ys	CATC	CATA
	alu GAG	ter X	AAAA
	550  1ys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys on GCA ACA ACA ACA CAC ACA CACA CACA CACA	570 ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter GCC GAG GAG AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA	CT AA
10	a E	500	CTGT
	ala OCT	3 4	دلادة
	ele GCT	<b>•</b> 55	۷
	age 21	ala GCT	.WAG
	550 \$50 \$70	580 gln CA	:CTC1
15	\$ 2	ser AGT	30110
	met ATG	ور ورک	7117
	2 F2	al GCT	11011
	85	val GTT	CTTT
20	* \$	3 E	ATCT
	36	1ys	ATTC
	58	* *	1130 TT
	29	99,	AAA
0.5	\$ <u>\$</u>	55	Q D
25	Sto thr	570 glu	<b>1</b> 50
	÷ 3	phe ala	<b>§</b>
			××
	) ) )	567 r eys c 160	ž Š
30	his lys cac aag	c thr	ter ter TCACAATAA
	⊋ 5 * <	s glu G GAG	A C.
	val lys h	p lys T AG	TACC
	5 C C	P esp	AGCC
	u leu IG CTC	a asp	ter ter Catctcagcctaccatgagaataagaaa
35	531 glu GAG	561 ala CCT	5

TCATTTTGCCTCTTTTGTCTGTGCTTCAATTAATAAAAATGGAAAGAATGTAA..... 20 .....AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5	(170)	(192)	(350)	(440)	(330)	(620)	(710)	3001
	20 1ys AAA	50 818 GCA	80 Leu CTT	110 Dro CCA	140 try TAT	170 qln CAA	200 0ys TGT	230 glu GAA (300)
	asn phe AAT TTC	phe TTT	thr	asn AAC	leu TTA	169 100 100	1 ys	ala GCA
	asn AAT	thr alu phe ACT CAA TTT	ala thr GCA ACT	ase GAC	tyr	168 cys TGT	leu CTC	phe TTT
10	g Ju GAA	th <b>r</b> ACT	val GTT	asp GAT	lys AAA	2 8 8	ara Aga	ala CAG
	GA A	glu val GAA GTA	thr ACA		1ys AAA	chr ACA	CAG	ala CCT
	lys asp leu qly glu glu asn phe AAA GAT TTG GGA GAA GAA TTC	E &	55 55 T	leu gin his lys TIG CAA CAC AAA	phe leu lys TTT TTG AAA	Phe TTT	lys AAA	1ys AAA
	leu TTG	asn AAT	75 lys leu cys AAA TTA TGC	c d a	phe	ala CCT	818 CCC	pro CCC
	asp	val GTG	asp lys CAC AAA	lev TTG	thr ACA	e) a	ser TCT	phe pro
15	1ys AAA	leu TTA	98 <b>p</b> <b>G</b> AC	phe TTC	alu GAG	1ys	ser TCG	ACA ACA
	10 erg phe lys CCC TTT AAA	40 val lys leu val asn GTA AAA TTA GTG AAT	<del>,</del> 69	101 103 105	ag ag	tyr TAT		gln arq CAG AGA
	٥٠ وتق م		70 phe III	100 GAA	130 asn qiu qiu thr phe leu lys lys tyr AAT GAA GAG ACA TTT TTG AAA AAA TAC	160 erg tyr lys ele ele ACG TAT AAA GCT GCT	190 qly lys ala GGC AAG GCT	
	elo his GCT CAT	phe alu asp his TTT GAA CAT CAT	thr leu ACC CTT	asn	åsb GAC	1 78	₹ 99 260	220 ala arg leu ser GCT CGC CTG AGC
00		g g	thr	874 AGA	hls CAT	ala CCT	골종	<b>်</b> သ
20	glu val GAG GTT	<u>1</u> 80	bis CAT	91. 000	ala phe GCT TTT	ohe TTT	asp GAT	ala CCT
	g tu CAC	phe TTT	Jeu CTT	pro CCT		phe TTC	ere SSS	
	ser Agt	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	ser TCA	gin glu CAA GAA	th. ACT	leu	leu CTT	ela GCA
	lys AAG	34 0ys 1GT	1ys AAA	g P CAA	12% 0y3 TGC	leu CTC	ale G&	lys ala trp ala val AAA GCA TGG GCA GTA
25	h1s CAC	gln CAG	asp GAC	1ys AAA	met ATG	glu	leu asp CTC GAT	ala SCA
	ala GCA	gln	62 0ys 1G1	ala CCA	val GTG	970	leu CTC	lys AAA
	sap GAT	Jeu CTT	60 glu asn GAA AAT	93 93 161	asp	818 CC	1ys AAG	phe TTC
		30 ala gin tyr leu GCT CAG TAT CTT		90 cys 1GC	120 val CTT	150 tyr TAT	180 Pro CCA	210 ala GCT
		ala gin CCT CAG	ela GCT	asp GAC	ar Gac	tyr phe TAC TTT	leu leu. 27G 77G	ara AGA
30		ala CCT	36. 1CA	ala CCT	pro CCA	tyr TAC	leu CTG	<b>28</b>
		phe TTT	glu GAG	glu met GAA ATG	arg AGA	pro CCT	177 973 TCC	91. 86.4
		91a CCC	esp CAT	glu GAA	val GTG	his cat	် ရော	phe TTT
		11e ATT	val ala asp GTT GCT GAT	tyr gly glu met TAT GCT GAA ATG	Jeu 77G	arg ASA	177 asp lys ala ala oys GAT AAA GCT GCC TGC	1ys AA
35		1eu	val GTT	tyr TAT	ere co	å S	lys AAA	gla CAA
		leu val leu ile ala phe TTG GTG TTG ATT GCC TTT	51 53 lys the cys val AAA ACA TGT GTT	glu thr GAA ACC	leu pro arg leu val CTC CCC CGA TTG GTG	tle ele arg erg his pro ATT CCC ACA ACA CAT CCT		leu gln lys phe CTC CAA AAA TTT
		leu TTG	thr		leu CTC	11e Att	ala CCT	ser AGT
		21 ala CCC	17.8 A	81 8rg CGT	111 asn AAC	ale GAA	171 818 CCT	201 ala GCC

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 leu CTT	290 11e ATT	320 ala CCT	350 ala ccc	380 leu CCT	\$10 879 CGT	440 his cat	&70 9er AGT (	500 1ys AAA (	530 va) CTT (
	asp CAC	289 cys 1CC	tyr TAT	leu CTT	Pro CCT	va1 CTT	Lys AAA		bro CCC	
	ara ala AGC CCC	h Is CAC	asn	40 A	1ys AAA		438 cys TCT	oro val CCA GTA	val p	ala leu GCA CTT
10	A C C	ser Icr	1ys AAA	leu CTG	phe ITT	leu leu CTG TTA	437 cys 1GI		tyr Y	
	asp asp CAT CAC	glu lys GAA AAA	316 cys TGC	leu CTG	olo CAA	# T# 2)25	Iys AAA	lys thr AAA ACG	thr t	aln thr CAA ACT
	asb CAT		val GTT	leu CTG	asb CAT	asn	ACC	gg G	alu thr GAA ACA	X & G
	ala CCT	pro leu leu CCT CTG TTG	aso CAT	val	phe TTC	CAG	41° 500 500	his c	esp CAT O	oln lle lys lys CAA ATC AAG AAA
	253 cys TGT	Jeu CTG	1ys AAG	val GTC	val GTG	phe TTC	val GTG	leu h		le 1 TC A
15	leu glu CTT CAA		ser AGT	ser TCT	1 ys	1ys	1ys	val I	glu vel CAA GTC	gin ile CAA ATC
		1ys AAA	glu GAA	tyr TAC	ီး ဗင	tyr TAC	414 665		leu c CTG (	
	250 asp leu CAT CTG	280 q1u CAA	310 val CTT	340 835 CAT	370 tyr TAT	6 <del>2</del> 6	430 leu qly lys vəl CTA GCA AAA GTG	860 461 leu cys TTA IGT	490 ala 1 GCT 0	520 glu arq GAG AGA
	asp CAT	279 cys TGT	phe TTT	pro CCT	8 8 55 155	9. 53 7. 53				
20	91y 669	278 Cys TGC	880 CAT	n is	ale GAA	Jeu CTT	ser ang asn TCA AGA AAC	leu asn gln CTG AAC CAG	phe ser ITT ICA	qlu lys CAG AAG
20	his Cat	ole GA	ala CCT	arg AGC	h is CAT	eln CAG	ser TCA	leu a		ser q TCT 0
	246 cys 760	leu lys CTG AAG	ala CCT	arg Aga	pro	alu GAG		val GTC	pro cys CCA TGC	leu ser CTT TCT
	2&5 6%3 16C		leu TTA	ela GCA	asp CAT	phe	glu vel GAG GTC	val ,		514 cys thr leu TCC ACA CTT
	thr glu ACG CAA	ser lys AGT AAA	ser TCA	tyr TAT	ala GCA			ser 700	arg arg AGG CGA	514 cys t TCC A
25		ser AGT	pro	<b>5</b> 83	ala CCT	glu leu GAG CTT	leu val CTT GTA	leu :	asn a	
	h Is CAC	ger TCC	leu TTG	tyr TAT	112 CCC	392 cys TCT	the ACT (	tyr I	val a	asp 11e Cat Ata
	val	11e ATC	asa GAC	leu TTG	361 cys TGT	esn AAT	er g	as b CAC	leu v TTG (	ala cca o
	240 173 AAA	270 ser TCG	300 e1e ect	330 phe TTT	360 cys 160	8 4 4 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	thr thr ACT	\$50 glu GAA	480 ser leu TCC TTG	510 hís s cat c
	thr ACC	asp CAT	pro CCT	gly met GCC ATG	1ys AAG	1ys AAA	ser TCA	ala GCA	. age	phe h
30	Jeu CTT	gla CA	met ATG		g Ga	ILE 198 ATC AAA	val	948 973 TGT (	thr g	thr p
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	glu asn CAA AAT	glu val glu asn asp glu GAA GTG GAA AAT GAT GAG	phe leu TTC TTG	thr leu ACT CTA		gla CAA	200		
	th <b>r</b> ACA	g1u GA	asp CAT	phe TTC	thr	asn	. 679 000	met 1 ATG (	476 t cys c TGC 1	asn ala glu thr phe AAT GCT GAA ACA TTC
	val CTG	265 cys TCT	glu val glu asn GAA GTG GAA AAT	val GTC	thr Acc	gla		arg met AGA ATG	28	glu thr GAA ACA
35	leu TTA	tyr 11e TAT ATC	g lu GAA	asp CAT	9 A	Pro CCT	lys AA	k A	بر بر ان بر	18 g CT 0
-	1ys AAG		val GTG	1ys AAG		P CAC	lys lys aac aaa	glu ala lys GAA CCA AAA	al t	phe asn ala TTT AAT CCT
	ser TCC	261 ala lys GCC AAG	g lu GAA	ala GCA	ACA ACA	ole SAA	the 1 ACC /	glu a	2 G	he a TTA
	231 val CTT	261 ala GCC	291 a1a ccc	321 glu ala lys asp val GAG GCA AAG GAT GTC	151 1ys AAG	381 val glu glu pro gln asn leu GTG GAA GAG CCT CAG AAT TTA	411 tyr thr lys lys val TAC ACC AAG AAA GTA	pro g	471 asp arg val thr lys cys cys GAC AGA GTC ACC AAA TGC TGC	501 glu phe CAC TTT

5	559 560 0ys 1ys 1GC AAG (1790)	ter TTAAAAG (1883)	TETTTAA (2002)	
10	558 559 5 e vel glu lys cys cys l T GTA GAG AAG TGC 1GC /	ala leu qly leu ter GCC TTA GCC TTA TAA CATCACATTTAAAAG (1883)	GTCTAAAAACATAAATI	
15	950 asp phe ala ala phe val glu lys o	580 gin ala ala leu qiy leu ter CAA GCT GCC TTA GCC TTA TAA	GTGTAAAGCCAACACCCT	
20	540 Iys ala thr Iys glu gin leu Iys ele val met asp e AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT (	phe ala glu glu qly lys lys leu val ala ala ser or coc cac cac ccr AAA AAA CTT GTT GCT GCA AGT	CTGTTTTTTTCGTTG	20AA (2078
25	lys glu gin leu ly AAA GAG CAA CTG A <sup>7</sup>	giu qiy iys iys ie Cac cct aaa aaa Ci	gatcaaaaccttattcat	atggaagaatctaa
30		567 570 thr cys phe ala glu ACC TCC TTT GCC GAG	tet Ataagagaagaaatgaa	gtgettcaattaataaaa
35	531 glu leu val Iys his Iys pro GAG CTC GTG AAA CAC AAG CCC	561 ala asp asp lys glu thr cys GCT GAC GAT AAG GAG ACC TGC	ter ter Catcicagectaccataggaataggaaagaaaatgaagaicaaagcttaticateigitititettitiggtgtaaagccaacaccetettaaaaagataattietttaa (2002)	TCATTTIGCCICTTITCICIGIGCTICAATTAAAAAATGGAAAGAATCTAA 20AA (2078)

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

	(30)
5	AGC
	ag II
	leu ohe CTC TTT
	ohe I
	2 I.
10	-10 11 C
	-10 ser leu leu phe l TCC CTT CTT TTT C
	phe 11e
15	tlu p
	7 44 T
	b p r o i lys trp val tlu b i AAG TGG GTA ACC
	% \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \
	Tet 3
20	-18 HEL STITTETETETETETETETETETETETETETETETETETE
	<b>₩</b>
	S CV
25	ננכע
23	CAAG
	רבה המא
	TCTT TCTT
	TTTTCTCTTCTGI
30	67 P P P P P P P P P P P P P P P P P P P
	919
	601 -6 p r - arg 91y val
35	tyr TAT
35	<b>61 9 1</b>
	100

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

5	(071)	(360)	(350)	(940)	(330)	(420)	(017)	(360)
	20 17s	50 818 600	80 leu CTT	110 pro CCA	140 try TAT	170 GAA	200 cys 7GT	230 glu GAA
	ohe TTC	phe TTT	thr	asn AAC	leu TTA	169 6ys 1GC	1 ys	#]# CCA
	asn	ale CAA	ela GCA	ass CAC	TAC	168 679 1GT	leu CTC	phe TTT
	910 GAA	thr ACT		asp GAT	lys AAA	gg ag	AGA AGA	alu Gac
10	5 & &	val GTA	thr val	1ys AAA	1 ys	ACA	ela CAG	ala GCT
	5 5	₹.8	25 998 100	hís CAC	leu TTG	phe TTT	1 ys	Lys AAA
	leu 176	asn AAT	lys leu AAA TTA	gln CAA	phe TTT	ala phe GCT TTT	ala lys GCC AAA	pro CCC
	esp CAT	val CTG	75 89p lys leu oys GAC AAA TTA TGC	leu 17G	ACA ACA	ele GCT		ohe TTT
15	phe lys TTT AAA	leu TTA	asp GAC	phe TTC	alu GAG	<b>8 8</b>	ser ser ICG ICI	arg AGA
	phe TTT	1ys AAA	91y CGA	100 873 351	91c	Eyr Tat		gln CAG
	ე გ.გ ე	40 val GTA	70 phe TTT	5 5 8 8	130 asn AAT	160 erg AGG	190 173 AGC	220 ser AGC
	his CAT	F S	leu CTT	asn AAT	esp GAC	1ys AAA	<del>,</del> 8	leu CTG
	ola GCT	asp CAT	his thr CAT ACC	P AS	h s CAT	ale GCT	a)u GAA	وتة رون
20	val CTT	SA A	h is CAT	91y 666	phe	phe	asp CAT	ela CCT
	ser glu AGT GAG	phe TTT	ser leu TCA CTT	Pro CCT	al a GCT	leu phe CTT TTC	န် သ	ala val CCA GTA
	ser AGT	S S S S S S S S S S S S S S S S S S S		og Pro	thr Act	leu CTT	leu	ala GCA
	1ys AAG	34 0ys 1CT	asp lys GAC AAA	gla	124 cys 760	leu CTC	91u GAA	<b>է</b> ը
25	his CAC	gln	asp CAC	1ys AAA	met ATG	38	asp CAT	ala CCA
	910 GCA	gla	62 0ys TGT	ala GCA	val GTG	ဋိပ္ပ	leu CTC	lys AAA
	i asp GAT	leu CTT	asn	91 693 1GT	asp GAT	ela OCC	1ys AAG	phe TTC
	- క్రి	30 tyr TAT	91.6 SA 5	98 93 165	120 val GTT	150 tyr TAT	180 Pro CCA	210 ala GCT
	arg CGT	gln CAG	ala CCT	asp GAC	g)u GAG	phe TTT	leu TTG	arg AGA
30	phe TTT	ala ccT	ser TCA	ala	rg CCA	tyr TAC	leu CTG	g gr
	و دتا دت	phe TTT	glu GAG	met ATG	AGA AGA	pro CCT	177 9ys 1CC	
	91 <i>y</i> GGT	#1# ccc	asp	glu met GAA ATG	leu val arg TTG GTG AGA	hås pro CAT CCT	177 ale eye GCC TGC	phe gly TTT CCA
	9-9 ACC	leu 11e TTC ATT	ala CCT	gly GGT	leu TTG	P	ala GCT	gin lys CAA AAA
		leu TTG	val	tyr TAT	5 S	9 4 4	1ys AAA	gla
35		21 818 1eu val 1eu 11e GCC 7TG GTG 7TG ATT	53 0ys 1GT	thr		ສຸເສ ເວ	asp GAT	leu CTC
		leu 1TG	thr	glu GAA	leu CTC	11e ATT	ela GCT	Ser
		21 ala GCC	2 2 8 8 8	81 arg CGT	111 asn AAC	£ 58	171 ala GCT	201 818 800

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 1eu CTT	290 11e ATT	320 ala err	350 ala ccc	380 leu CCT	#10 erg	440 h1s cat	470 9er AGT (	500 1 ys AAA (	
<b>-</b>	est CAC	289 cys 100	êye Tat		bro CCT	val CTT			Pro 1 CCC A	520 ser glu lys glu arg gln fle lys lys gln thr ala leu val TCT GAG AAG GAG AGA GAA ACT GCA CTT GTT
	ele SC	h lo CAC	AAC	arg AGA	Bys AAA	leu TTA	438 Cys lys ICI AAA	<b>Dro val</b> CCA <b>G</b> TA	val p GTT C	2 S
10	8 <del>2</del> 8		lys Aaa	leu CTG	phe	leu CTG	637 6 cys TGT 1	thr p	tyr v Tac G	2 S
	asp arg	plu lys GAA AAA	316 cys lys TCC AAA	leu leu CTG CTG	g d GAA	als GC O		A X	thr ty ACA T	= = = = = = = = = = = = = = = = = = =
	SSP CAT	leu leu glu lys ser CTG TTG GAA AAA TCC	val G17	ser val val leu leu leu arg leu TCT GTC GTG CTG CTG AGA CTT	val phe asp glu GTG TTC CAT CAA	asn a	ser lys ACC AAA	leu hís glu lys TTG CAT GAG AAA	alu thr GAA ACA	\$ & G
	ele CT	leu leu CTG TTG	880 GAT	val leu GTG CTG	phe TTC	cyc /		Is q	asp a	2 S
	elu cys CAA TGT	leu CTG	lys AAG	val GTC	cre cre	phe o	val aly GTG GGC	leu hís TTG CAT	val a	ie 13 TC AV
15		Pro CCT	ser A <b>GT</b>	ser	1 ys	lys IAA		val l	gľu v CAA G	5 S
	250 leu leu CTG CTT	280 qlu lya GAA AAA	glu ser lys asp val CAA AGT AAC GAT GTT	ty TAC	ala iya vai phe GCC AAA GTG TTC	tyr lys Tac aaa	oly Bys Cca Aaa	461 Cys v TGT G	leu g CTG G	6 S
		280 414 643	310 val CTT	340 asp tyr CAT TAC	370 tyr	400 glu tyr lys GAG TAC AAA	436 leu c CTA (	\$60 & leu c TTA T	490 ala leu glu val asp GCT CTG GAA GTG GAT	510 ala asp ile cys thr leu ser glu lys glu arg CCA CAT ATA ICC ACA CTT ICT CAG AAG GAG AGA
	asp CAT	279 cy8 TGT	asp phe CAT TTF	pro	369 cys TGC	78	\$36 asn leu AAC CTA	gln 1 CAG T	aer a TCA G	8 5 G
20	hla gly CAT GGA	278 0ys 1GC		his CAT	동동	phe alu ain Icu aly TTT GAG CAG CTT GGA		esn g AAC C		le 1
20	h is CAT	g Jr GAA	ala GCT	arg AGC	his CAT	CAG.	ser arg TCA AGA	leu e CTG A	8 % T	e 5
	23.6 29.5 160	lys leu lys AAA CTG AAG	ela CCT	arg ACA	pro CCT	2 2 2		val J Grc c	pro cys phe CCA TGC TTT	er s
	265 glu eys GAA TGC	leu CTG	ser leu TCA TTA	ala GCA	asp CAT	phe (	glu val GAG GTC	val v Gro o	arg p CGA C	518 ala asp ile cys thr leu CCA CAT ATA TCC ACA CTT
		A A	ser TCA	tyr Tat	ele GCA	leu crr		3er v 7CC G	arg a	51& cys tl TCC A
25	his thr CAC ACG	oer ser ICC ACT	org CCT	5 8 8	ala GCT	2 5 2 5	leu v	leu s CTA 1	asn arg AAC AGG	S le c
	240 Iyo vol his the Aaa GTC CAC ACG		leu TTG	tyr Tat	ala ecc	392 cys TCT	thr leu val ACT CTT GTA	tyr I	val a	sp 1 AT A
	240 1yo vol Aaa CTC	270 ser fle TCG ATC	esp GAC	leu 1TC	361 cys TGT		Pro	asp c	480 ser leu val TCC TTG GTG	18 8 CA G
•			300 ala cc1	330 phe 111	360 9ys 1GC	390 gin asn CAA AAT	\$20 thr pro ACT CCA	\$50 91u 6 GAA 0	480 ser 1 TCC 1	510 hís a cat c
	leu the CTT ACC	asp CAT	pro CCT	gly met GGC ATG	1ys AAG	1 ys	ser TCA	2 S	glu s CAA 7	510 phe h13 TTC CAT
30		gln CAA	met AT <b>C</b>	91y 000	91c GAG	11e ATC	val	446 0ys 1 TCT (	ر د م و	
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	glu asn CAA AAT	91e Se	leu TTG	leu CTA	ssn leu fle lys AAT TTA ATC AAA	pro gin val	pro cys ala	cys t	he t TC A
•	chr.		asp GAT	phe TTC	thr	AAT	. 50	met I	476 4 Cy3 C	7 2
	vai CTG	265 0ys TCT	asn AAT	val GTC	thr Acc	c AG	val GTA	<b>5</b> 43	\$ %	1. A A
35	leu TTA	tyr 11e Tat atc	glu GA	asp CAT	936	65	% ¥ .	. ys 6	بر کر ۳ ک	la CT C
	1ys AAG	tyr Tat	val GTG	1ys AAG	tyr TAT	GAG	AG A	18 1 CA A	al t	sn a AT G
		261 ala lys tyr lle GCC AAG TAT ATC	291 ala glu val glu asn GCC GAA GTG GAA AAT	321 glu ala lys asp val phe GAG GCA AAG GAT GTC TTC	351 1ys thr tyr glu thr thr AAG ACA TAT GAA ACC ACT	glu glu pro gln GAA GAG CCT CAG	tit tyr thr lys lys val pro TAC ACC AAG AAA GTA CCC	glu ala lys arg GAA GCA AAA AGA	ភ្ ខ្ម > ប	phe asn ala glu thr phe thr III AAI GCI GAA ACA IIC ACC
	231 val GTT	261 ala CCC	291 a1a GCC	321 91u GAG	351 173	381 val CTC	411 tyr 1	441 010 001	471 asp arg val thr lys cys thr GAC AGA GTC ACC AAA TGC TGC ACA	501 glu phe asn ala glu thr phe thr GAG ITT AAT GCT GAA ACA TTC ACC

	_	_	_
5	560 1ys AAG (1790)	ter Taa Catcacatttaaaag (1883)	ter ter Catetcagcetaccatgagaataagagaaagaaagaatcaaagcettattcatetetttttegttggtgtaaagccaacacctgtctaaaaacataaatttctttaa (2002) Tcattttgcetcttttetetggttcaattaataaaaatggaaagaatctaa 20aa (2078)
-	560 1ys AAG (	86	TAA (
	8	ter	ICTT
	558 559 560 cys cys lys TCC TCC AAG	CATI	ATTI
	γ AG α α	ATCA	ATA
10	550 asp phe ala ala phe val glu lys cys cys CAT TTC CCT GCT TTT CTA CAC AAC TGC TGC	er A	IAAAC
	al g	leu t TTA 1	TAA/
	y ₽ = C	ala leu qly leu ter GCC TTA GCC TTA TAA	7670
	18 P.	leu qly TTA GCC	ACCC
	4 E	1 ete	CAAC
15	် ပွဲ (၁)	@ E	AGC
	550 asp pt CAT TI	580 gln ala CAA GCT	IGTA
	550 P asp IT CAT	7 5 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	TCC
	ຶ ວີ ມຸບ	a se A AC	TCG1
00	I me	a aj CC	1111 •
20	5 T	1 of 2	ווונ
	9 9 1	8 5	rcrr, 20
	1 ys	S CTI	ATC
	leu CTG	₹ 8	ATTC TAA.
25	g P	1ys AA	CCTT AATC
25	glu gin leu lys sis val met ssp e	41.y	AAAG
	1ye	916 646	PATC/
	540 1ys ala thr 1ys glu gln AAG GCA ACA AAA GAG CAA	570 phe ala glu glu qly lys lys leu val ala ala ser TTT GCC GAG-GAG GGT AAA AAA CTT GTT GCT GCA AGT	GAAC
	ala GCA	ala ငင	AAA1 ATA
30	1ys AAG	phe TTT	AAGA Atta
			989 11C8
	Age	thr	ter ter Itcacaataa Itcicigigo
	AC AC	alu GAG	ar Gaga CTCT
	8 A	lys AG	CATTT
35	leu val lys his lys pro CTC GTG AAA CAC AAG CCC	567 asp asp lys glu thr cys GAC GAT AAG GAG ACC TGC	ter ter Catcicagcctaccatgagaataagagaaatgaagatcaaagcttaticatctgititicgitgg Tcatitigcctcttitctgigcttcaattaataaaaatggaaagaaictaa 20aa (2078)
	ee ,	Sep of	7. CAQ
	531 glv 1 GAG C	561 818 6 CCT (	SATCI
	~ -0	N 8 O	J P

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

5	(30)	(170)	(560)	(350)	(U\$\$)	(330)	(620)	(710)	(300)
	Ber AGC	20 178 AAA	50 818 878	88 184 CTT	110 Pro CCA	140 Ery TAT	170 qin CAA	200 cys TGT	230 91u GAA
	0 he	phe TTC	alu phe CAA III	sla thr CCA ACT	asa	leu TTA	169 cys 100	1ys AAC	618 GCA
	leu CTC	AAT			ese CAC	tyr TAC	168 Cys TGT	leu CTC	Phe TT
10	g F o vol thu phe lie ser leu leu phe leu phe GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT	alu alu GAA GAA	thr	75 leu oys thr val TTA TCC ACA GTT	ohe leu qin his iys aso TTC TTG CAA CAC AAA GAT	phe leu lys lys TTT TTG AAA AAA	ale GAS	r a	a tu GAG
	leu CTT	5 &9 6¥	glu val GAA GTA	en Aca	1ys AAA	1ys AAA		al CAG	ala Gr.T
	- 10 대한 기업	leu aly TTG GGA	g g	2 % 10 %	qin his Caa cac	leu TTG	phe TTT	Lys AAA	1 ys AAA
	ser TCC	leu TTG	asn AAT	leu TTA	<b>g In</b> CAA	phe	ele ect	ရှင် ၁၃၁	و <u>د</u> در
	phe Blo TTT ATT	lys osb leu aly Aaa gat ttg gga	leu val ITA GTG	asp lys CAC AAA	leu TTG	glu thr GAG ACA	ela GCT	ser	phe TTT
15	pho TTT	phe lys TTT AAA	Je TTA	70 phe gly asp lys TTT CGA GAC AAA	ohe TTC	ale GAG	1 ys	ser 1CG	erg AGA
	g r o voltlu GTA ACC	phe TTT	40 his val 198 cat gta aaa	4 833	100 101 alv eys CAA TCC	as GAS	tyr Tat	ala GCT	220 ser gin arg AGC CAG AGA
	o vol GTA	ර දැය සිනි	% Vel GTA			130 asn AAT		130 133 AAG	220 ser gin AGC CAG
	466	his Cat	r is Cat	thr lew ACC CTT	esn AAT	asa GAC	ala lys GCT AAA	alu aly CAA CCC	leu CTG
20	8 yo	010 <b>GCT</b>	CAT CAT		AGA GA	his CAT	ala CCT		5 8
20	-18 Kat ATG	glu vol CAC CTT	<b>3</b> 89	leu hlis CTT CAT	9 te 666	ala phe CCT TTT	ohe TTT	88 <b>5</b>	ele GCT
			phe TTT	leu CTT	pro CCT		phe TTC	arg asp CGC GAT	vel GTA
		i oop olo Mis Iyo Ger Gat CCA CAC AAG AGT	34 cys pro IGT CCA	ser	gle GAA	thr	leu leu CTC CTT	leu CTT	trp ala val TGG GCA GTA
		lyo Ag	φ cys 1G1	Tys AAA	c AA	124 078 1GC	leu CTC		trp ala TGG GCA
25		olo his CCA CAC	gln CAG	dese CAC	1 ys	met ATG	glu GAA	asp CAT	3 A A
		၀ <b>။</b> ၀ <b>၄၄</b>	gla CAG	62 cys TGT	olo GCA	val	5 5 5 5 5 5 5 5	leu CTC	1 ys
		1 ბან <b>GAT</b>	30 tyr leu TAT CTT	60 glu asn GAA AAT	90 91 eys eys IGC IGI	asp GAT	ရူ	180 pro lys cca AAG	phe TTC
		- 493 693	30 tyr TAT	95 e		120 val GTT	150 tyr TAT	180 CCA	210 ala GCT
		phe org	gln	ala GCT	asp GAC	g]u GAG	phe TTT	leu leu CTG TTG	8 4 A C A
30		ghe TT	ols GCT	ser ICA	ala SCT	r S	tyr TAC	leu CTG	glu arg GAA AGA
35		gly vol cct ctc	phe TTT	9 6 6	glu met GAA ATG	P P P	pro CCT	177 eys 100	91 <b>,</b> 66 A
			ala OCC	asp CAT	38	pro arg leu val arg CCC CGA TTG GTG AGA	ala arg arg his pro CCC AGA AGA CAT CCT	177 ele eys GCC TGC	phe TTT
		27.9 AGC	11e ATT	ele CCT	ely ccr	leu TTG	arg AGA	ala GCT	Lys AAA
		2. 100	ોલ્ય 176	val CTT	tyr Tat	5 A	AGA AG	1ys AAA	를 A S
		olo tyr oer org cer tar tee aes	21 alo leu val leu lle CCC ITG GTG TTG ATT	53 thr bys val ACA TGT GTT	glu thr GAA ACC	222	ala OCC	esp GAT	leu CTC
		o10 <b>cc1</b>	Je 176	ACA A	<b>3</b> 6	leu CTC	116	al a GCT	ser AGT
		Dor	ole OCC	17s AAA	919 CGT	111 asn AAC	181 CAA /	171 e1s CCT	201 ala ser leu gin lys phe gly CCC AGT CTC CAA AAA TTT CCA

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 1eu CTT	290 11e Att	320 ele CCT	350 818 050	380 1eu CCT	&10 ara CGT	840 NIS CAT	470 ger AGT	Sno 1 ys AAA	530 val GTT
	ess CAC	289 cys 70C	tyr Tat	leu	Pro	val	173 AAA	val GTA	ورر (درر	leu C7T
	9 Je 9 CG	his CAC	asn	ara AGA	1ys AAA	leu TTA	438 cys TGT	<b>در</b> کا	val	ele GCA
10	a rg AGG	ser TCC	1 ys	leu CTG	phe TTT	leu CTG	437 cys TGT		tyr TAC	thr ACT
	685 GAC	1ys AAA	316 cys 1GC	leu CTG	GA S	ala OCC	ser lys AGC AAA	elu lys GAG AAA	thr ACA	£ &
	ola asp CCT CAT	3 €	val	ser val val leu leu leu ard TCT GTC GTG CTG CTG AGA	asp CAT	esn		ege Gae		ain lie lys lys ain thr Caa atc aac aaa caa act
	ور وريا	pro leu leu CCT CTG TTG	asp CAT	val GTG	lys val phe AAA GTG TTC	gln CAG	4 <del>1</del> 6	leu his ITG CAT	asp CAT	1ys AAG
	253 Cys 7C7	leu CTG	ser lys AGT AAG	val GTC	val GTG	lys phe AAA TTC	val		val	11e ATC
15	g ic GAA		glu ser lys Caa act aac		ala iya val GCC AAA GTG	Py8	lys AAA	val	<b>3</b> 89	
	250 leu leu CTG CTT	280 qlu lys Gaa aaa		340 asp tyr GAT TAC	ရှိ ရှိ ပိပ်	400 glu tyr lyg GAG TAC AAA	430 leu qly CTA GGA	460 461 leu cys val TTA TGT GTG	890 ala leu CCT CTG	arg AGA
		280 91u GAA	310 vel CTT	340 889 GAT	369 370 qlu cys tyr GAA TGC TAT	91u 98c	430 1eu CTA	\$60 Jeu		520 leu ser glu lys glu arq CTT TCT CAC AAG CAG AGA
	809 GA₹	278 279 eys eys TCC TCT	phe TTT	his pro CAT CCT	369 698 160	leu aly CTT GGA	asın AAC	CAG	phe ser TTF TCA	lys AAG
20	9 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	278 978 100	esp CAT		gje GAA	leu CTT	AGA AGA	leu asn gln CTG AAC CAG		91u GAG
60	Mgo CAF	lys glu AAG GAA	919	tyr glu tyr ala arg arg TAT GAA TAT GCA AGA AGG	S I S	gla CAG	ser TCA	Jeu CTG	pro eya CCA TGC	514 oys thr leu ser TCC ACA CTT TCT
	245 246 giu oyo oyo gaa tac tac	1ye	ala GCT	arg AGA	5 TO	g gr	val	val	CCA	leu CTT
	245 gilu oyo gaa tgc	lys leu AAA CTG	ser leu ICA ITA	ela GCA	els esp. GCA GAT	leu phe CTT TTT	glu	leu ser val CTA TCC GTG	65 690	514 88p lle gys thr CAT ATA TCC ACA
				glu tyr GAA TAT	8 S	leu CTT	thr leu val ACT CTT GFA	ser TCC	arg AGG	
25	hio the Cac acg	Ser ACT	leu pro TTG CCT	28	ala GCT	GAG CAG	leu	leu CTA	asn AAC	11e Ata
		Ser TCC			#1# CCC	392 cys TGT	thr	tyr	val GTG	asp CAT
	280 Iyo vol Aaa GTC	11e	SAC D	leu TTG	360 361 oys cys TGC TGT	asn AAT	420 thr pro ACT CCA	450 glu asp GAA GAC	leu TTG	510 his ala asp ile CAT CCA GAT ATA
		270 ser TCG	300	330 phe 111		3% GA			480 ser TCC	
	Ehr.	88p	pro	gly met GGC ATG	glu lys GAG AAG	173 848	ser TCA	448 cys ale ICT CCA	glu GAA	thr phe ACC TTC
30	leu CTT	gln	met : ATG			11e ATC	val		thr ACA	
35	. 20g	glu asn GAA AAT	oge .	leu TTG	tyr glu thr thr leu TAT GAA ACC ACT CTA	leu TTA	gla	520	477 0ys TGC	glu thr phe GAA ACA TTC
	S AC		vel glu esn esp GTG GAA AAT GAT	phe TTC	thr ACT	asn AAT	. 50	ATG	\$76 9ys 1GC	AG AG
	vol	265 11e cys ATC TGT	asını A	val	thr ACC	gla	val GTA	AGA	1ys	2 g
	let TT/	. 11e ATC		asp GAT	gle SA	glu pro GAG CCT	1ya	AAA	thr	asn ala AAT GCT
	ser by her vol the TCC aag tta GTG aca	tyr TAT	glu val GAA GTG	1ys	tyr TAT	glu GAG	the lys lys vol pro ACC AAG AAA GTA CCC	glu bla lys arg met Caa CCa aaa aca atg	val GTC	asn AAT
		1 AAG		321 glu ala lys asp val phe GAG GCA AAG GAT GTC TTC	the ACA	ale s	thr Acc	48	ACA	501 glu phe CAG TTT
	239 Vol GTT	261 a1a ccc	291 818 GCC	321 91u GAG	351 1ys AAG	381 val GTG	411 tyr TAC	441 pro	471 asp GAC	501 914 6A6

5		(1790)	leu ter TTA TAA CATCACATTTAAAAG (1883)	ter Catcicagcctaccatgagaataagaaaagaaatgaagatcaaagcttaticatctgttititctttttggtggtaggccaacaccctgtctaaaaaacataaatttctttaa (20112)
		560 175 AAG	AG	¥.
		559 100 100	ter ITAA	ICTT
		1ys cys c	ACAT	AATT
10		lys AAG	CATC	CATA
10		glu GAG	ter TAA	AAAA
		phe sis sis phe vol giu TC GCT GCT TTT GTA CAG	ala alo leu qly leu ter GCT GCC TTA GGC TTA TAA	CTAA
		<b>ph</b> e 111	leu qly TTA GCC	CTGT
		ala ccr	leu TTA	CACC
15	•	ele GCT	၁၂၀	CCA
		ghe TTC	ala CCT	TAAAC
		550 asp asp car	570 ala glu glu qiy iys iys leu val ola ala ser gin ala alo 3 GCC GAG GAG GG7 AAA AAA CTT GTT GC7 GCA AGT CAA GCT GCC 3	SCTGI
		es TA	ser AGT	cerre
00		540 thr lyo glu gin leu lys slo vol met ACA AAA GAG CAA CTG AAA GCT GTT ATG	ele GCA	rrr
20		val	9 607	TTCT
		910	val GT	CTTT
	•	1 yes	leu CTT	ATCT
		18c	lys AAA	ATTC
25		gla CAA	1 ya	8CT
		970	<b>4</b> 50 50 50 50 50 50 50 50 50 50 50 50 50	AAAA
		AAC A	2 de 2 de 2	SA C
		S&O thr	570 91u	2
		ele ccA	a 55	\$¥¥
30		1 ys	phe TTT	PAAAC
		ord a	567 678 700	a WGAC
		his lys cac aag	c Acc	ter ter Tcacaataa
	٠	a A	s glu G GG	ter VTGA(
35		val lys	lys r AAG	'ACC'
		C 49	es s	locc1
		ı u Jeu G CTC	a asp	וכזכי
		531 91u 6A6	561 ala CCT	ક

TCATTITGCCTCTTTTCTCTGCTTCAATTAAAAAATGGAAAGAATCTAA..... 20 .....AA (2078)

- 10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
- 11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
  - 13. A DNA transfer vector according to claim 12, which is a plasmid.
- 14. A DNA transfer vector according to claim 13,10 wherein the plasmid is pBR322 or YEp6.
  - 15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
  - 16. A DNA transfer vector according to any of
- 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
  - 17. A vector or process according to claim 16, wherein the bacterium or yeast is  $\underline{E}$ .  $\underline{coli}$  or  $\underline{Saccharomyces}$   $\underline{cerevisiae}$ .

1/1 Mpo () (3658) (3613) (18513) (18513) © ⊗ ۵. 586 ferm TAA ۵. ف Restriction Endonuclease Map of Human Serum Albumin cDNA Clones 493/4 Tag ! 419/0 450/1 Ninc II Mbo II ۵. ه <u>ک</u> ھ pMA36 382 Sbo # ڪ الا Kilobases 325/6 Mbe II ဓာ Pot i (3611) Hpå II (3548) 182/3 Tag l ณ์ Hp2 II (3548) Pst I (3511) æ, pHA208 ಣ 57 Hid I Pst 1 Mbo II Mbo II (3611) 16/7 31 બં Hpa II (3658)